Fourth Edition

LINEAR ALGEBRA
AND ITS APPLICATIONS

Gilbert Strang

39. From m independent measurements $b_1,\dots,b_m$ of your pulse rate, weighted by $w_1,\dots,w_m$, what is the weighted average that replaces equation (9)? It is the best estimate when the statistical variances are $\sigma_i^2 \equiv 1/w_i^2$.

40. If $W = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$, find the W-inner product of $x = (2,3)$ and $y = (1,1)$, and the W-length of x. What line of vectors is W-perpendicular to y?

41. Find the weighted least-squares solution $\widehat{x}_W$ to $Ax = b$:

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \qquad b = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \qquad W = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Check that the projection $A\widehat{x}_W$ is still perpendicular (in the W-inner product!) to the error $b - A\widehat{x}_W$.

42. (a) Suppose you guess your professor's age, making errors $e = -2, -1, 5$ with probabilities $\frac{1}{2}, \frac{1}{4}, \frac{1}{4}$. Check that the expected error $E(e)$ is zero and find the variance $E(e^2)$.

(b) If the professor guesses too (or tries to remember), making errors -1, 0, 1 with probabilities $\frac{1}{8}, \frac{6}{8}, \frac{1}{8}$, what weights w_1 and w_2 give the reliability of your guess and the professor's guess?

3.4 Orthogonal Bases and Gram-Schmidt

In an orthogonal basis, every vector is perpendicular to every other vector. The coordinate axes are mutually orthogonal. That is just about optimal, and the one possible improvement is easy: Divide each vector by its length, to make it a *unit vector*. That changes an ***orthogonal*** basis into an ***orthonormal*** basis of q's:

3P The vectors $q_1,\dots,q_n$ are ***orthonormal*** if

$$q_i^{\mathrm{T}} q_j = \begin{cases} 0 & \text{whenever } i \neq j, \quad \text{giving the orthogonality;} \\ 1 & \text{whenever } i = j, \quad \text{giving the normalization.} \end{cases}$$

A matrix with orthonormal columns will be called Q.

The most important example is the *standard basis*. For the x-y plane, the best-known axes $e_1 = (1,0)$ and $e_2 = (0,1)$ are not only perpendicular but horizontal and vertical. Q is the 2 by 2 identity matrix. In n dimensions the standard basis $e_1,\dots,e_n$ again consists of *the columns of* $Q = I$:

Standard basis
$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \cdots, \quad e_n = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}.$$

That is not the only orthonormal basis! We can rotate the axes without changing the right angles at which they meet. These rotation matrices will be examples of Q.

If we have a subspace of $\mathbf{R}^n$, the standard vectors e_i might not lie in that subspace. But the subspace always has an orthonormal basis, and it can be constructed in a simple way out of any basis whatsoever. This construction, which converts a skewed set of axes into a perpendicular set, is known as ***Gram-Schmidt orthogonalization***.

To summarize, the three topics basic to this section are:

1. The definition and properties of orthogonal matrices Q.
2. The solution of $Qx = b$, either n by n or rectangular (least squares).
3. The Gram-Schmidt process and its interpretation as a new factorization $A = QR$.

Orthogonal Matrices

3Q If Q (square or rectangular) has orthonormal columns, then $Q^{\mathrm{T}}Q = I$:

Orthonormal columns
$$\begin{bmatrix} — & q_1^{\mathrm{T}} & — \\ — & q_2^{\mathrm{T}} & — \\ & \vdots & \\ — & q_n^{\mathrm{T}} & — \end{bmatrix} \begin{bmatrix} | & | & & | \\ q_1 & q_2 & \cdots & q_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdot & 0 \\ 0 & 1 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & 1 \end{bmatrix} = I. \tag{1}$$

An orthogonal matrix is a square matrix with orthonormal columns.[2] Then Q^{T} is Q^{-1}. For square orthogonal matrices, ***the transpose is the inverse***.

When row i of Q^{T} multiplies column j of Q, the result is $q_j^{\mathrm{T}}q_j = 0$. On the diagonal where $i = j$, we have $q_i^{\mathrm{T}}q_i = 1$. That is the normalization to unit vectors of length 1.

Note that $Q^{\mathrm{T}}Q = I$ even if Q is rectangular. But then Q^{T} is only a left-inverse.

Example 1.

$$Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}, \qquad Q^{\mathrm{T}} = Q^{-1} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}.$$

[2] Orthonormal matrix would have been a better name, but it is too late to change. Also, there is no accepted word for a rectangular matrix with orthonormal columns. We still write Q, but we won't call it an "orthogonal matrix" unless it is square.

Q rotates every vector through the angle θ, and Q^{T} rotates it back through $-\theta$. The columns are clearly orthogonal, and they are orthonormal because $sin^2\theta + cos^2\theta = 1$. The matrix Q^{T} is just as much an orthogonal matrix as Q.

Example 2. Any permutation matrix P is an orthogonal matrix. The columns are certainly unit vectors and certainly orthogonal—because the 1 appears in a different place in each column: The transpose is the inverse.

$$\text{If} \quad P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \quad \text{then} \quad P^{-1} = P^{\mathrm{T}} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

An anti-diagonal P, with $P_{13} = P_{22} = P_{31} = I$, takes the x-y-z axes into the z-y-x axes—a "right-handed" system into a "left-handed" system. So we were wrong if we suggested that every orthogonal Q represents a rotation. *A reflection is also allowed.* $P = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$ reflects every point (x,y) into (y,x), its mirror image across the 45° line. Geometrically, an orthogonal Q is the product of a rotation and a reflection.

There does remain one property that is shared by rotations and reflections, and in fact by every orthogonal matrix. It is not shared by projections, which are not orthogonal or even invertible. Projections reduce the length of a vector, whereas orthogonal matrices have a property that is the most important and most characteristic of all:

3R Multiplication by any Q preserves lengths:

$$\textbf{Lengths unchanged} \qquad \|Qx\| = \|x\| \quad \text{for every vector } x. \qquad (2)$$

It also preserves inner products and angles, since $(Qx)^{\mathrm{T}}(Qy) = x^{\mathrm{T}}Q^{\mathrm{T}}Qy = x^{\mathrm{T}}y$.

The preservation of lengths comes directly from $Q^{\mathrm{T}}Q = I$:

$$\|Qx\|^2 = \|x\|^2 \quad \text{because} \quad (Qx)^{\mathrm{T}}(Qx) = x^{\mathrm{T}}Q^{\mathrm{T}}Qx = x^{\mathrm{T}}x. \qquad (3)$$

All inner products and lengths are preserved, when the space is rotated or reflected.

We come now to the calculation that uses the special property $Q^{\mathrm{T}} = Q^{-1}$. If we have a basis, then any vector is a combination of the basis vectors. This is exceptionally simple for an orthonormal basis, which will be a key idea behind Fourier series. The problem is *to find the coefficients of the basis vectors*:

$$\boxed{\textbf{\textit{Write } } b \textbf{\textit{ as a combination}} \quad b = x_1q_1 + x_2q_2 + \cdots + x_nq_n.}$$

To compute x_1 there is a neat trick. *Multiply both sides of the equation by* q_1^{T}. On the left-hand side is $q_1^{\mathrm{T}}b$. On the right-hand side all terms disappear (because $q_1^{\mathrm{T}}q_j = 0$) except the first term. We are left with

$$q_1^{\mathrm{T}}b = x_1q_1^{\mathrm{T}}q_1.$$

Since $q_1^{\mathrm{T}}q_1 = 1$, we have found $x_1 = q_1^{\mathrm{T}}b$. Similarly the second coefficient is $x_2 = q_2^{\mathrm{T}}b$; that term survives when we multiply by q_2^{T}. The other terms die of orthogonality. Each piece of b has a simple formula, and recombining the pieces gives back b:

$$\textbf{\textit{Every vector } } b \textbf{\textit{ is equal to }} (q_1^{\mathrm{T}}b)q_1 + (q_2^{\mathrm{T}}b)q_2 + \cdots + (q_n^{\mathrm{T}}b)q_n. \tag{4}$$

I can't resist putting this orthonormal basis into a square matrix Q. The vector equation $x_1q_1 + \cdots + x_nq_n = b$ is identical to $Qx = b$. (The columns of Q multiply the components of x.) Its solution is $x = Q^{-1}b$. But since $Q^{-1} = Q^{\mathrm{T}}$—this is where orthonormality enters—the solution is also $x = Q^{\mathrm{T}}b$:

$$x = Q^{\mathrm{T}}b = \begin{bmatrix} — & q_1^{\mathrm{T}} & — \\ & \vdots & \\ — & q_n^{\mathrm{T}} & — \end{bmatrix} \begin{bmatrix} \\ b \\ \\ \end{bmatrix} = \begin{bmatrix} q_1^{\mathrm{T}}b \\ \vdots \\ q_n^{\mathrm{T}}b \end{bmatrix} \tag{5}$$

The components of x are the inner products $q_i^{\mathrm{T}}b$, as in equation (4).

The matrix form also shows what happens when the columns are *not* orthonormal. Expressing b as a combination $x_1a_1 + \cdots + x_na_n$ is the same as solving $Ax = b$. The basis vectors go into the columns of A. In that case we need A^{-1}, which takes work. In the orthonormal case we only need Q^{T}.

***Remark* 1.** The ratio $a^{\mathrm{T}}b/a^{\mathrm{T}}a$ appeared earlier, when we projected b onto a line. Here a is q_1, the denominator is 1, and the projection is $(q_1^{\mathrm{T}}b)q_1$. Thus we have a new interpretation for formula (4): *Every vector b is the sum of its one-dimensional projections onto the lines through the q's.*

Since those projections are orthogonal, Pythagoras should still be correct. The square of the hypotenuse should still be the sum of squares of the components:

$$\|b\|^2 = (q_1^{\mathrm{T}}b)^2 + (q_2^{\mathrm{T}}b)^2 + \cdots + (q_n^{\mathrm{T}}b)^2 \quad \text{which is} \quad \|Q^{\mathrm{T}}b\|^2. \tag{6}$$

Remark* 2.** Since $Q^{\mathrm{T}} = Q^{-1}$, we also have $QQ^{\mathrm{T}} = I$. When Q comes before Q^{T}, multiplication takes the inner products of the *rows* of Q. (For $Q^{\mathrm{T}}Q$ it was the columns.) Since the result is again the identity matrix, we come to a surprising conclusion: ***The rows of a square matrix are orthonormal whenever the columns are. The rows point in completely different directions from the columns, and I don't see geometrically why they are forced to be orthonormal—but they are.

Orthonormal columns
Orthonormal rows

$$Q = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & 0 & -2/\sqrt{6} \\ 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6} \end{bmatrix}.$$

Rectangular Matrices with Orthogonal Columns

This chapter is about $Ax = b$, when A is not necessarily square. For $Qx = b$ we now admit the same possibility—there may be more rows than columns. The n orthonormal vectors q_i in the columns of Q have $m > n$ components. Then Q is an m by n matrix and we cannot expect to solve $Qx = b$ exactly. *We solve it by least squares.*

If there is any justice, orthonormal columns should make the problem simple. It worked for square matrices, and now it will work for rectangular matrices. The key is to notice that *we still have $Q^{\mathrm{T}}Q = I$*. So Q^{T} is still the ***left-inverse*** of Q.

For least squares that is all we need. The normal equations came from multiplying $Ax = b$ by the transpose matrix, to give $A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}b$. Now the normal equations are $Q^{\mathrm{T}}Q = Q^{\mathrm{T}}b$. But $Q^{\mathrm{T}}Q$ is the identity matrix! Therefore $\widehat{x} = Q^{\mathrm{T}}b$, whether Q is square and $\widehat{x}$ is an exact solution, or Q is rectangular and we need least squares.

3S If Q has orthonormal columns, the least-squares problem becomes easy: rectangular system with no solution for most b.

$$\begin{aligned} Qx &= b && \text{rectangular system with no solution for most } b. \\ Q^{\mathrm{T}}Q\widehat{x} &= Q^{\mathrm{T}}b && \text{normal equation for the best } \widehat{x}\text{—in which } Q^{\mathrm{T}}Q = I. \\ \widehat{x} &= Q^{\mathrm{T}}b && \widehat{x}_i \text{ is } q_i^{\mathrm{T}}b. \\ p &= Q\widehat{x} && \text{the projection of } b \text{ is } (q_1^{\mathrm{T}}b)q_1 + \cdots + (q_n^{\mathrm{T}}b)q_n. \\ p &= QQ^{\mathrm{T}}b && \text{the projection matrix is } P = QQ^{\mathrm{T}}. \end{aligned}$$

The last formulas are like $p = A\widehat{x}$ and $P = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$. When the columns are orthonormal, the "cross-product matrix" $A^{\mathrm{T}}A$ becomes $Q^{\mathrm{T}}Q = I$. The hard part of least squares disappears when vectors are orthonormal. The projections onto the axes are uncoupled, and p is the sum $p = (q_1^{\mathrm{T}}b)q_1 + \cdots + (q_n^{\mathrm{T}}b)q_n$.

We emphasize that those projections do not reconstruct b. In the square case $m = n$, they did. In the rectangular case $m > n$, they don't. They give the projection p and not the original vector b—which is all we can expect when there are more equations than unknowns, and the q's are no longer a basis. The projection matrix is usually $A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$, and here it simplifies to

$$P = Q(Q^{\mathrm{T}}Q)^{-1}Q^{\mathrm{T}} \qquad \text{or} \qquad P = QQ^{\mathrm{T}}. \tag{7}$$

Notice that $Q^{\mathrm{T}}Q$ is the n by n identity matrix, whereas QQ^{T} is an m by m projection P. It is the identity matrix on the columns of Q (P leaves them alone), But QQ^{T} is the zero matrix on the orthogonal complement (the nullspace of Q^{T}).

Example 3. The following case is simple but typical. Suppose we project a point $b = (x, y, z)$ onto the x-y plane. Its projection is $p = (x, y, 0)$, and this is the sum of the separate projections onto the x- and y-axes:

$$q_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad (q_1^{\mathrm{T}}b)q_1 = \begin{bmatrix} x \\ 0 \\ 0 \end{bmatrix}; \qquad q_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad (q_2^{\mathrm{T}}b)q_2 = \begin{bmatrix} 0 \\ y \\ 0 \end{bmatrix}.$$

The overall projection matrix is

$$P = q_1q_1^{\mathrm{T}} + q_2q_2^{\mathrm{T}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \text{and} \quad P\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x \\ y \\ 0 \end{bmatrix}.$$

Projection onto a plane = sum of projections onto orthonormal q_1 ***and*** q_2.

Example 4. When the measurement times average to zero, fitting a straight line leads to orthogonal columns. Take $t_1 = -3$, $t_2 = 0$, and $t_3 = 3$. Then the attempt to fit $y = C + Dt$ leads to three equations in two unknowns:

$$\begin{array}{rcl} C + Dt_1 &=& y_1 \\ C + Dt_2 &=& y_2 \\ C + Dt_3 &=& y_3 \end{array}, \qquad \text{or} \qquad \begin{bmatrix} 1 & -3 \\ 1 & 0 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}.$$

The columns $(1,1,1)$ *and* $(-3,0,3)$ *are orthogonal.* We can project y separately onto each column, and the best coefficients $\widehat{C}$ and $\widehat{D}$ can be found separately:

$$\widehat{C} = \frac{\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} y_1 & y_2 & y_3 \end{bmatrix}^{\mathrm{T}}}{1^2+1^2+1^2}, \qquad \widehat{D} = \frac{\begin{bmatrix} -3 & 0 & 3 \end{bmatrix}\begin{bmatrix} y_1 & y_2 & y_3 \end{bmatrix}^{\mathrm{T}}}{(-3)^2+0^2+3^2}.$$

Notice that $\widehat{C} = (y_1+y_2+y_3)/3$ is the *mean* of the data. $\widehat{C}$ gives the best fit by a horizontal line, whereas $\widehat{D}t$ is the best fit by a straight line through the origin. *The columns are orthogonal, so the sum of these two separate pieces is the best fit by any straight line whatsoever*. The columns are not unit vectors, so $\widehat{C}$ and $\widehat{D}$ have the length squared in the denominator.

Orthogonal columns are so much better that it is worth changing to that case. if the average of the observation times is not zero—it is $\bar{t} = (t_1 + \cdots + t_m)/m$—then the time origin can be shifted by $\bar{t}$. Instead of $y = C + Dt$ we work with $y = c + d(t - \bar{t})$. The best line is the same! As in the example, we find

$$\begin{aligned} \widehat{c} &= \frac{\begin{bmatrix} 1 & \cdots & 1 \end{bmatrix}\begin{bmatrix} y_1 & \cdots & y_m \end{bmatrix}^{\mathrm{T}}}{1^2+1^2+\cdots+1^2} = \frac{y_1+\cdots+y_m}{m} \\ \widehat{d} &= \frac{\begin{bmatrix} (t_1-\bar{t}) & \cdots & (t_m-\bar{t}) \end{bmatrix}\begin{bmatrix} y_1 & \cdots & y_m \end{bmatrix}^{\mathrm{T}}}{(t_1-\bar{t})^2+\cdots+(t_m-\bar{t})^2} = \frac{\sum(t_i-\bar{t})y_i}{\sum(t_i-\bar{t})^2}. \end{aligned} \tag{8}$$

The best $\widehat{c}$ is the mean, and we also get a convenient formula for $\widehat{d}$. The earlier $A^{\mathrm{T}}A$ had the off-diagonal entries $\sum t_i$, and shifting the time by $\bar{t}$ made these entries zero. This shift is an example of the Gram-Schmidt process, ***which orthogonalizes the situation in advance***.

Orthogonal matrices are crucial to numerical linear algebra, because they introduce no instability. While lengths stay the same, roundoff is under control. Orthogonalizing vectors has become an essential technique. Probably it comes second only to elimination. And it leads to a factorization $A = QR$ that is nearly as famous as $A = LU$.

The Gram-Schmidt Process

Suppose you are given three independent vectors a, b, c. If they are orthonormal, life is easy. To project a vector v onto the first one, you compute $(a^{\mathrm{T}}v)a$. To project the same vector v onto the plane of the first two, you just add $(a^{\mathrm{T}}v)a+(b^{\mathrm{T}}v)b$. To project onto the span of a, b, c, you add three projections. All calculations require only the inner products $a^{\mathrm{T}}v$, $b^{\mathrm{T}}v$, and $c^{\mathrm{T}}v$. But to make this true, we are forced to say, "***If*** they are orthonormal." Now we propose to find a way to ***make*** them orthonormal.

The method is simple. We are given a, b, c and we want q_1, q_2, q_3. There is no problem with q_1: it can go in the direction of a. We divide by the length, so that $q_1 = a/\|a\|$ is a unit vector. The real problem begins with q_2—which has to be orthogonal to q_1. If the second vector b has any component in the direction of q_1 (which is the direction of a), ***that component has to be subtracted***:

$$\textbf{Second vector} \qquad B = b-(q_1^{\mathrm{T}}b)q_1 \quad \text{and} \quad q_2 = B/\|B\|. \tag{9}$$

B is orthogonal to q_1. It is the part of b that goes in a new direction, and not in the a. In Figure 3.10, B is perpendicular to q_1. It sets the direction for q_2.

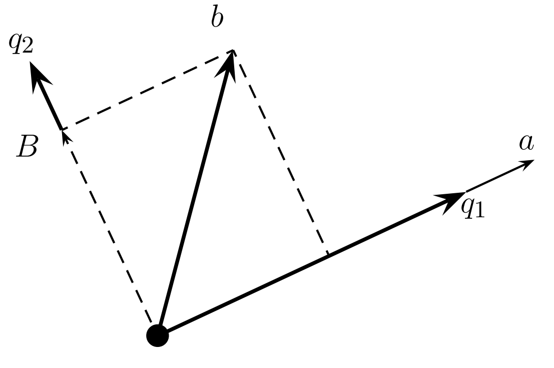

Figure 3.10: The q_i component of b is removed; a and B normalized to q_1 and q_2.

At this point q_1 and q_2 are set. The third orthogonal direction starts with c. It will not be in the plane of q_1 and q_2, which is the plane of a and b. However, it may have a component in that plane, and that has to be subtracted. (If the result is $C=0$, this signals that a, b, c were not independent in the first place) What is left is the component C we want, the part that is in a new direction perpendicular to the plane:

$$\textbf{Third vector} \qquad C = c-(q_1^{\mathrm{T}}c)q_1-(q_2^{\mathrm{T}}c)q_2 \quad \text{and} \quad q_3 = C/\|C\|. \tag{10}$$

This is the one idea of the whole Gram-Schmidt process, ***to subtract from every new vector its components in the directions that are already settled***. That idea is used over and over again.[3] When there is a fourth vector, we subtract away its components in the directions of q_1, q_2, q_3.

[3]If Gram thought of it first, what was left for Schmidt?

Example 5. Gram-Schmidt Suppose the independent vectors are a, b, c:

$$a = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \qquad b = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \qquad c = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}.$$

To find q_1, make the first vector into a unit vector: $q_1 = a/\sqrt{2}$. To find q_2, subtract from the second vector its component in the first direction:

$$B = b - (q_1^{\mathrm{T}} b) q_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \frac{1}{\sqrt{2}} \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}.$$

The normalized q_2 is B divided by its length, to produce a unit vector:

$$q_2 = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ -1/\sqrt{2} \end{bmatrix}.$$

To find q_3, subtract from c its components along q_1 and q_2:

$$\begin{aligned} C &= c - (q_1^{\mathrm{T}} c) q_1 - (q_2^{\mathrm{T}} c) q_2 \\ &= \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} - \sqrt{2} \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix} - \sqrt{2} \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ -1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}. \end{aligned}$$

This is already a unit vector, so it is q_3. I went to desperate lengths to cut down the number of square roots (the painful part of Gram-Schmidt). The result is a set of orthonormal vectors q_1, q_2, q_3, which go into the columns of an orthogonal matrix Q:

$$\textbf{Orthonormal basis} \qquad Q = \begin{bmatrix} & & \\ q_1 & q_2 & q_3 \\ & & \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \end{bmatrix}.$$

3T The Gram-Schmidt process starts with independent vectors $a_1, \ldots, a_n$ and ends with orthonormal vectors $q_1, \ldots, q_n$. At step j it subtracts from a_j its components in the directions $q_1, \ldots, q_{j-1}$ that are already settled:

$$A_j = a_j - (q_1^{\mathrm{T}} a_j) q_1 - \cdots - (q_{j-1}^{\mathrm{T}} a_j) q_{j-1}. \tag{11}$$

Then q_j is the unit vector $A_j/\|A_j\|$.

Remark on the calculations I think it is easier to compute the orthogonal a, B, C, without forcing their lengths to equal one. Then square roots enter only at the end, when dividing by those lengths. The example above would have the same B and C, without using square roots. Notice the $\frac{1}{2}$ from $a^{\mathrm{T}}b/a^{\mathrm{T}}a$ instead of $\frac{1}{\sqrt{2}}$ from $q^{\mathrm{T}}b$:

$$B = \begin{bmatrix}1\\0\\0\end{bmatrix} - \frac{1}{2}\begin{bmatrix}1\\0\\1\end{bmatrix} \quad \text{and then} \quad C = \begin{bmatrix}2\\1\\0\end{bmatrix} - \begin{bmatrix}1\\0\\1\end{bmatrix} - 2\begin{bmatrix}\frac{1}{2}\\0\\-\frac{1}{2}\end{bmatrix}.$$

The Factorization $A = QR$

We started with a matrix A, whose columns were a, b, c. We ended with a matrix Q, whose columns are q_1, q_2, q_3. What is the relation between those matrices? The matrices A and Q are m by n when the n vectors are in m-dimensional space, and there has to be a third matrix that connects them.

The idea is to write the a's as combinations of the q's. The vector b in Figure 3.10 is a combination of the orthonormal q_1 and q_2, and we know what combination it is:

$$b = (q_1^{\mathrm{T}}b)q_1 + (q_2^{\mathrm{T}}b)q_2.$$

Every vector in the plane is the sum of its q_1 and q_2 components. Similarly c is the sum of its q_1, q_2, q_3 components: $c = (q_1^{\mathrm{T}}c)q_1 + (q_2^{\mathrm{T}}c)q_2 + (q_3^{\mathrm{T}}c)q_3$. If we express that in matrix form we have ***the new factorization*** $A = QR$:

$$\textbf{\textit{QR} factors} \qquad A = \begin{bmatrix} & & \\ a & b & c \\ & & \end{bmatrix} = \begin{bmatrix} & & \\ q_1 & q_2 & q_3 \\ & & \end{bmatrix} \begin{bmatrix} q_1^{\mathrm{T}}a & q_1^{\mathrm{T}}b & q_1^{\mathrm{T}}c \\ & q_2^{\mathrm{T}}b & q_2^{\mathrm{T}}c \\ & & q_3^{\mathrm{T}}c \end{bmatrix} = QR \qquad (12)$$

Notice the zeros in the last matrix! R is *upper triangular* because of the way Gram-Schmidt was done. The first vectors a and q_1 fell on the same line. Then q_1, q_2 were in the same plane as a, b. The third vectors c and q_3 were not involved until step 3.

The QR factorization is like $A = LU$, except that the first factor Q has orthonormal columns. The second factor is called R, because the nonzeros are to the *right* of the diagonal (and the letter U is already taken). The off-diagonal entries of R are the numbers $q_1^{\mathrm{T}}b = 1/\sqrt{2}$ and $q_1^{\mathrm{T}}c = q_2^{\mathrm{T}}c = \sqrt{2}$, found above. The whole factorization is

$$A = \begin{bmatrix}1 & 1 & 2\\0 & 0 & 1\\1 & 0 & 0\end{bmatrix} = \begin{bmatrix}1/\sqrt{2} & 1/\sqrt{2} & 0\\0 & 0 & 1\\1/\sqrt{2} & -1/\sqrt{2} & 0\end{bmatrix} \begin{bmatrix}\sqrt{2} & 1/\sqrt{2} & \sqrt{2}\\ & 1/\sqrt{2} & \sqrt{2}\\ & & 1\end{bmatrix} = QR.$$

You see the lengths of a, B, C on the diagonal of R. The orthonormal vectors q_1, q_2, q_3, which are the whole object of orthogonalization, are in the first factor Q.

Maybe QR is not as beautiful as LU (because of the square roots). Both factorizations are vitally important to the theory of linear algebra, and absolutely central to the calculations. If LU is Hertz, then QR is Avis.

The entries $r_{ij} = q_i^{\mathrm{T}} a_j$ appear in formula (11), when $\|A_j\| q_j$ is substituted for A_j:

$$a_j = (q_1^{\mathrm{T}} a_j) q_1 + \cdots + (q_{j-1}^{\mathrm{T}} a_j) q_{j-1} + \|A_j\| q_j = Q \text{ times column } j \text{ of } R. \tag{13}$$

3U Every m by n matrix with independent columns can be factored into $A = QR$. The columns of Q are orthonormal, and R is upper triangular and invertible. When $m = n$ and all matrices are square, Q becomes an orthogonal matrix.

I must not forget the main point of orthogonalization. It simplifies the least-squares problem $Ax = b$. The normal equations are still correct, but $A^{\mathrm{T}}A$ becomes easier:

$$A^{\mathrm{T}}A = R^{\mathrm{T}}Q^{\mathrm{T}}QR = R^{\mathrm{T}}R. \tag{14}$$

The fundamental equation $A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}b$ simplifies to a triangular system:

$$R^{\mathrm{T}}R\widehat{x} = R^{\mathrm{T}}Q^{\mathrm{T}}b \qquad \text{or} \qquad R\widehat{x} = Q^{\mathrm{T}}b. \tag{15}$$

Instead of solving $QRx = b$, which can't be done, we solve $R\widehat{x} = Q^{\mathrm{T}}b$ which is just back-substitution because R is triangular. The real cost is the mn^2 operations of Gram-Schmidt, which are needed to find Q and R in the first place.

The same idea of orthogonality applies to functions, The sines and cosines are orthogonal; the powers $1, x, x^2$ are not. When $f(x)$ is written as a combination of sines and cosines, that is a ***Fourier series***. Each term is a projection onto a line—the line in function space containing multiples of $\cos nx$ or $\sin nx$. It is completely parallel to the vector case, and very important. And finally we have a job for Schmidt: To orthogonalize the powers of x and produce the Legendre polynomials.

Function Spaces and Fourier Series

This is a brief and optional section, but it has a number of good intentions:

1. to introduce the most famous infinite-dimensional vector space (*Hilbert space*);
2. to extend the ideas of length and inner product from vectors v to functions $f(x)$:
3. to recognize the Fourier series as a sum of one-dimensional projections (the orthogonal "columns" are the sines and cosines);
4. to apply Gram-Schmidt orthogonalization to the polynomials $1, x, x^2, \ldots$; and
5. to find the best approximation to $f(x)$ by a straight line.

We will try to follow this outline, which opens up a range of new applications for linear algebra, in a systematic way.

1. Hilbert Space. After studying $\mathbf{R}^n$, it is natural to think of the space $\mathbf{R}^\infty$. It contains all vectors $v = (v_1, v_2, v_3, \ldots)$ with an infinite sequence of components. This space is actually too big when there is no control on the size of components v_j. A much better idea is to keep the familiar definition of length, using a sum of squares, and *to include only those vectors that have a* ***finite length***:

$$\textbf{Length squared} \qquad \|v\|^2 = v_1^2 + v_2^2 + v_3^2 + \cdots \tag{16}$$

The infinite series must converge to a finite sum. This leaves $(1, \frac{1}{2}, \frac{1}{3}, \ldots)$ but not $(1, 1, 1, \ldots)$. Vectors with finite length can be added ($\|v+w\| \le \|v\| + \|w\|$) and multiplied by scalars, so they form a vector space. It is the celebrated ***Hilbert space***.

Hilbert space is the natural way to let the number of dimensions become infinite, and at the same time to keep the geometry of ordinary Euclidean space. Ellipses become infinite-dimensional ellipsoids, and perpendicular lines are recognized exactly as before. The vectors v and w are orthogonal when their inner product is zero:

$$\textbf{Orthogonality} \qquad v^{\mathrm{T}}w = v_1w_1 + v_2w_2 + v_3w_3 + \cdots = 0.$$

This sum is guaranteed to converge, and for any two vectors it still obeys the Schwarz inequality $|v^{\mathrm{T}}w| \le \|v\|\|w\|$. The cosine, even in Hilbert space, is never larger than 1.

There is another remarkable thing about this space: It is found under a great many different disguises. Its "vectors" can turn into functions, which is the second point.

2. Lengths and Inner Products. Suppose $f(x) = \sin x$ on the interval $0 \le x \le 2\pi$. This f is like a vector with a whole continuum of components, the values of $\sin x$ along the whole interval. To find the length of such a vector, the usual rule of adding the squares of the components becomes impossible. This summation is replaced, in a natural and inevitable way, by *integration*:

$$\textbf{Length } \|f\| \textbf{ of function} \qquad \|f\|^2 = \int_0^{2\pi} (f(x))^2 dx = \int_0^{2\pi} (\sin x)^2 dx = \pi \tag{17}$$

Our Hilbert space has become a ***function space***. The vectors are functions, we have a way to measure their length, and the space contains all those functions that have a finite length—just as in equation (16). It does not contain the function $F(x) = 1/x$, because the integral of $1/x^2$ is infinite.

The same idea of replacing summation by integration produces the ***inner product of two functions***: If $f(x) = \sin x$ and $g(x) = \cos x$, then their inner product is

$$(f, g) = \int_0^{2\pi} f(x)g(x)dx = \int_0^{2\pi} \sin x \cos x dx = 0. \tag{18}$$

This is exactly like the vector inner product $f^{\mathrm{T}}g$. It is still related to the length by $(f, f) = \|f\|^2$. The Schwarz inequality is still satisfied: $|(f, g)| \le \|f\|\|g\|$. Of course, two functions like $\sin x$ and $\cos x$—whose inner product is zero—will be called orthogonal. They are even orthonormal after division by their length $\sqrt{\pi}$.

3. The ***Fourier series*** of a function is an expansion into sines and cosines:

$$f(x) = a_0 + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x + \cdots .$$

To compute a coefficient like b_1, *multiply* both sides by the corresponding function $\sin x$ and *integrate* from 0 to 2π. (The function $f(x)$ is given on that interval.) In other words, take the inner product of both sides with $\sin x$:

$$\int_0^{2\pi} f(x) \sin x dx = a_0 \int_0^{2\pi} \sin x dx + a_1 \int_0^{2\pi} \cos x \sin x dx + b_1 \int_0^{2\pi} (\sin x)^2 dx + \cdots .$$

On the right-hand side, every integral is zero except one—the one in which $\sin x$ multiplies itself. *The sines and cosines are mutually orthogonal* as in equation (18) Therefore b_1 is the left-hand side divided by that one nonzero integral:

$$b_1 = \frac{\int_0^{2\pi} f(x) \sin x dx}{\int_0^{2\pi} (\sin x)^2 dx} = \frac{(f, \sin x)}{(\sin x, \sin x)}.$$

The Fourier coefficient a_1 would have $\cos x$ in place of $\sin x$, and a_2 would use $\cos 2x$.

The whole point is to see the analogy with projections. The component of the vector b along the line spanned by a is $b^{\mathrm{T}}a/a^{\mathrm{T}}a$. A ***Fourier series is projecting*** $f(x)$ ***onto*** $\sin x$. Its component p in this direction is exactly $b_1 \sin x$.

The coefficient b_1 is the least squares solution of the inconsistent equation $b_1 \sin x = f(x)$. This brings $b_1 \sin x$ as close as possible to $f(x)$. All the terms in the series are projections onto a sine or cosine. Since the sines and cosines are orthogonal, *the Fourier series gives the coordinates of the "vector" $f(x)$ with respect to a set of* (infinitely many) *perpendicular axes*.

4. Gram-Schmidt for Functions. There are plenty of useful functions other than sines and cosines, and they are not always orthogonal. The simplest are the powers of x, and unfortunately there is no interval on which even 1 and x^2 are perpendicular. (Their inner product is always positive, because it is the integral of x^2.) Therefore the closest parabola to $f(x)$ is *not* the sum of its projections onto 1, x, and x^2. There will be a matrix like $(A^{\mathrm{T}}A)^{-1}$, and this coupling is given by the ill-conditioned **Hilbert matrix**. On the interval $0 \leq x \leq 1$,

$$A^{\mathrm{T}}A = \begin{bmatrix} (1,1) & (1,x) & (1,x^2) \\ (x,1) & (x,x) & (x,x^2) \\ (x^2,1) & (x^2,x) & (x^2,x^2) \end{bmatrix} = \begin{bmatrix} \int 1 & \int x & \int x^2 \\ \int x & \int x^2 & \int x^3 \\ \int x^2 & \int x^3 & \int x^4 \end{bmatrix} = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{bmatrix}.$$

This matrix has a large inverse, because the axes 1, x, x^2 are far from perpendicular. The situation becomes impossible if we add a few more axes. *It is virtually hopeless to solve* $A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}b$ *for the closest polynomial of degree ten.*

More precisely, it is hopeless to solve this by Gaussian elimination; every roundoff error would be amplified by more than 10^{13}. On the other hand, we cannot just give up; approximation by polynomials has to be possible. The right idea is to switch to orthogonal axes (by Gram-Schmidt). We look for combinations of 1, x, and x^2 that *are* orthogonal.

It is convenient to work with a symmetrically placed interval like $-1 \le x \le 1$, because this makes all the odd powers of x orthogonal to all the even powers:

$$(1,x) = \int_{-1}^{1} x\,dx = 0, \qquad (x,x^2) = \int_{-1}^{1} x^3\,dx = 0.$$

Therefore the Gram-Schmidt process can begin by accepting $v_1 = 1$ and $v_2 = x$ as the first two perpendicular axes. Since $(x,x^2) = 0$, it only has to correct the angle between 1 and x^2. By formula (10), the third orthogonal polynomial is

$$\textbf{Orthogonalize} \qquad v_3 = x^2 - \frac{(1,x^2)}{(1,1)}1 - \frac{(x,x^2)}{(x,x)}x = x^2 - \frac{\int_{-1}^{1} x^2\,dx}{\int_{-1}^{1} 1\,dx} = x^2 - \frac{1}{3}.$$

The polynomials constructed in this way are called the ***Legendre polynomials*** and they are orthogonal to each other over the interval $-1 \le x \le 1$.

$$\textbf{Check} \qquad \left(1, x^2 - \frac{1}{3}\right) = \int_{-1}^{1}\left(x^2 - \frac{1}{3}\right)dx = \left[\frac{x^3}{3} - \frac{x}{3}\right]_{-1}^{1} = 0.$$

The closest polynomial of degree ten is now computable, without disaster, by projecting onto each of the first 10 (or 11) Legendre polynomials.

5. Best Straight Line. Suppose we want to approximate $y = x^5$ by a straight line $C + Dx$ between $x = 0$ and $x = 1$. There are at least three ways of finding that line, and if you compare them the whole chapter might become clear!

1. Solve $[1 \;\; x]\begin{bmatrix} C \\ D \end{bmatrix} = x^5$ by least squares. The equation $A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}b$ is

$$\begin{bmatrix} (1,1) & (1,x) \\ (x,1) & (x,x) \end{bmatrix}\begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} (1,x^5) \\ (x,x^5) \end{bmatrix} \qquad \text{or} \qquad \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{bmatrix}\begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} \frac{1}{6} \\ \frac{1}{17} \end{bmatrix}.$$

2. Minimize $E^2 = \int_0^1 (x^5 - C - Dx)^2 dx = \frac{1}{11} - \frac{2}{6}C - \frac{2}{7}D + C^2 + CD + \frac{1}{3}D^2$. The derivatives with respect to C and D, after dividing by 2, bring back the normal equations of method 1 (and the solution is $\widehat{C} = \frac{1}{6} - \frac{5}{14}$, $\widehat{D} = \frac{5}{17}$):

$$-\frac{1}{6} + C + \frac{1}{2}D = 0 \qquad \text{and} \qquad -\frac{1}{7} + \frac{1}{2}C + \frac{1}{3}D = 0.$$

3. Apply Gram-Schmidt to replace x by $x - (1,x)/(1,1)$. That is $x - \frac{1}{2}$, which is orthogonal to 1. Now the one-dimensional projections add to the best line:

$$C + Dx = \frac{(x^5,1)}{(1,1)}1 + \frac{(x^5, x - \frac{1}{2})}{(x - \frac{1}{2}, x - \frac{1}{2})}(x - \tfrac{1}{2}) = \frac{1}{6} + \frac{5}{7}\left(x - \frac{1}{2}\right).$$

Problem Set 3.4

1. (a) Write the four equations for fitting $y = C + Dt$ to the data

$$\begin{array}{llllll} y=-4 & \text{at} & t=-2, & y=-3 & \text{at} & t=-1 \\ y=-1 & \text{at} & t=1, & y=0 & \text{at} & t=2. \end{array}$$

Show that the columns are orthogonal.

(b) Find the optimal straight line, draw its graph, and write E^2.

(c) Interpret the zero error in terms of the original system of four equations in two unknowns: The right-hand side $(-4,-3,-1,0)$ is in the ____ space.

2. Project $b = (0,3,0)$ onto each of the orthonormal vectors $a_1 = (\frac{2}{3}, \frac{2}{3}, -\frac{1}{3})$ and $a_2 = (-\frac{1}{3}, \frac{2}{3}, \frac{2}{3})$, and then find its projection p onto the plane of a_1 and a_2.

3. Find also the projection of $b = (0,3,0)$ onto $a_3 = (\frac{2}{3}, -\frac{1}{3}, \frac{2}{3})$, and add the three projections. Why is $P = a_1a_1^{\mathrm{T}} + a_2a_2^{\mathrm{T}} + a_3a_3^{\mathrm{T}}$ equal to I?

4. If Q_1 and Q_2 are orthogonal matrices, so that $Q^{\mathrm{T}}Q = I$, show that Q_1Q_2 is also orthogonal. If Q_1 is rotation through θ, and Q_2 is rotation through ϕ, what is Q_1Q_2? Can you find the trigonometric identities for $\sin(\theta+\phi)$ and $\cos(\theta+\phi)$ in the matrix multiplication Q_1Q_2?

5. If u is a unit vector, show that $Q = I - 2uu^{\mathrm{T}}$ is a symmetric orthogonal matrix. (It is a reflection, also known as a Householder transformation.) Compute Q when $u^{\mathrm{T}} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix}$.

6. Find a third column so that the matrix

$$Q = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{14} & \\ 1/\sqrt{3} & 2/\sqrt{14} & \\ 1/\sqrt{3} & -3/\sqrt{14} & \end{bmatrix}$$

is orthogonal. It must be a unit vector that is orthogonal to the other columns; how much freedom does this leave? Verify that the rows automatically become orthonormal at the same time.

7. Show, by forming $b^{\mathrm{T}}b$ directly, that Pythagoras's law holds for any combination $b = x_1q_1 + \cdots + x_nq_n$ of orthonormal vectors: $\|b\|^2 = x_1^2 + \cdots + x_n^2$. In matrix terms, $b = Qx$, so this again proves that lengths are preserved: $\|Qx\|^2 = \|x\|^2$.

8. Project the vector $b = (1,2)$ onto two vectors that are not orthogonal, $a_1 = (1,0)$ and $a_2 = (1,1)$. Show that, unlike the orthogonal case, the sum of the two one-dimensional projections does not equal b.

9. If the vectors q_1, q_2, q_3 are orthonormal, what combination of q_1 and q_2 is closest to q_3?

10. If q_1 and q_2 are the outputs from Gram-Schmidt, what were the possible input vectors a and b?

11. Show that an orthogonal matrix that is upper triangular must be diagonal.

12. What multiple of $a_1 = \left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]$ should be subtracted from $a2 = \left[\begin{smallmatrix}4\\0\end{smallmatrix}\right]$ to make the result orthogonal to a_1? Factor $\left[\begin{smallmatrix}1&4\\1&0\end{smallmatrix}\right]$ into QR with orthonormal vectors in Q.

13. Apply the Gram-Schmidt process to

$$a = \begin{bmatrix}0\\0\\1\end{bmatrix}, \qquad b = \begin{bmatrix}0\\1\\1\end{bmatrix}, \qquad c = \begin{bmatrix}1\\1\\1\end{bmatrix}$$

and write the result in the form $A = QR$.

14. From the nonorthogonal a, b, c, find orthonormal vectors q_1, q_2, q_3:

$$a = \begin{bmatrix}1\\1\\0\end{bmatrix}, \qquad b = \begin{bmatrix}1\\0\\1\end{bmatrix}, \qquad c = \begin{bmatrix}0\\1\\1\end{bmatrix}.$$

15. Find an orthonormal set q_1, q_2, q_3 for which q_1, q_2 span the column space of

$$A = \begin{bmatrix}1 & 1\\ 2 & -1\\ -2 & 4\end{bmatrix}.$$

Which fundamental subspace contains q_3? What is the least-squares solution of $Ax = b$ if $b = [1\ \ 2\ \ 7]^{\mathrm{T}}$?

16. Express the Gram-Schmidt orthogonalization of a_1, a_2 as $A = QR$:

$$a_1 = \begin{bmatrix}1\\2\\2\end{bmatrix}, \qquad a_2 = \begin{bmatrix}1\\3\\1\end{bmatrix}.$$

Given n vectors a_i with m components, what are the shapes of A, Q, and R?

17. With the same matrix A as in Problem 16, and with $b = [1\ \ 1\ \ 1]^{\mathrm{T}}$, use $A = QR$ to solve the least-squares problem $Ax = b$.

18. If $A = QR$, find a simple formula for the projection matrix P onto the column space of A.

19. Show that these *modified Gram-Schmidt* steps produce the same C as in equation (10):

$$C^* = c - (q_1^{\mathrm{T}} c) q_1 \qquad \text{and} \qquad C = C^* - (q_2^{\mathrm{T}} C^*) q_2.$$

This is much more stable, to subtract the projections one at a time.

20. In Hilbert space, find the length of the vector $v = (1/\sqrt{2}, 1/\sqrt{4}, 1/\sqrt{8}, \ldots)$ and the length of the function $f(x) = e^x$ (over the interval $0 \le x \le 1$). What is the inner product over this interval of e^x and e^{-x}?

21. What is the closest function $a\cos x + b\sin x$ to the function $f(x) = \sin 2x$ on the interval from $-\pi$ to π? What is the closest straight line $c + dx$?

22. By setting the derivative to zero, find the value of b_1 that minimizes

$$\|b_1 \sin x - \cos x\|^2 = \int_0^{2\pi} (b_1 \sin x - \cos x)^2 dx.$$

Compare with the Fourier coefficient b_1.

23. Find the Fourier coefficients a_0, a_1, b_1 of the step function $y(x)$, which equals 1 on the interval $0 \le x \le \pi$ and 0 on the remaining interval $\pi < x < 2\pi$:

$$a_0 = \frac{(y,1)}{(1,1)} \qquad a_1 = \frac{(y,\cos x)}{(\cos x, \cos x)} \qquad b_1 = \frac{(y, \sin x)}{(\sin x, \sin x)}.$$

24. Find the fourth Legendre polynomial. It is a cubic $x^3 + ax^2 + bx + c$ that is orthogonal to 1, x, and $x^2 - \frac{1}{3}$ over the interval $-1 \le x \le 1$.

25. What is the closest straight line to the parabola $y = x^2$ over $-1 \le x \le 1$?

26. In the Gram-Schmidt formula (10), verify that C is orthogonal to q_1 and q_2.

27. Find an orthonormal basis for the subspace spanned by $a_1 = (1,-1,0,0)$, $a_2 = (0,1,-1,0)$, $a_3 = (0,0,1,-1)$.

28. Apply Gram-Schmidt to $(1,-1,0)$, $(0,1,-1)$, and $(1,0,-1)$, to find an orthonormal basis on the plane $x_1 + x_2 + x_3 = 0$. What is the dimension of this subspace, and how many nonzero vectors come out of Gram-Schmidt?

29. (Recommended) Find orthogonal vectors A, B, C by Gram-Schmidt from a, b, c:

$$a = (1,-1,0,0) \qquad b = (0,1,-1,0) \qquad c = (0,0,1,-1).$$

A, B, C and a, b, c are bases for the vectors perpendicular to $d = (1,1,1,1)$.

30. If $A = QR$ then $A^{\mathrm{T}}A = R^{\mathrm{T}}R =$ ____ triangular times ____ triangular. *Gram-Schmidt on A corresponds to elimination on $A^{\mathrm{T}}A$.* Compare

$$A = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix} \qquad \text{with} \qquad A^{\mathrm{T}}A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}.$$

For $A^{\mathrm{T}}A$, the pivots are 2, $\frac{3}{2}$, $\frac{4}{3}$ and the multipliers are $-\frac{1}{2}$ and $-\frac{2}{3}$.

(a) Using those multipliers in A, show that column 1 of A and B = column 2 $-\frac{1}{2}$(column 1) and C = column 3 $-\frac{2}{3}$(column 2) are orthogonal.

(b) Check that $\|\text{column } 1\|^2 = 2$, $\|B\|^2 = \frac{3}{2}$, and $\|C\|^2 = \frac{4}{3}$, using the pivots.

31. True or false (give an example in either case):

(a) Q^{-1} is an orthogonal matrix when Q is an orthogonal matrix.

(b) If Q (3 by 2) has orthonormal columns then $\|Qx\|$ always equals $\|x\|$.

32. (a) Find a basis for the subspace $\mathbf{S}$ in $\mathbf{R}^4$ spanned by all solutions of

$$x_1 + x_2 + x_3 - x_4 = 0.$$

(b) Find a basis for the orthogonal complement $\mathbf{S}^\perp$.

(c) Find b_1 in $\mathbf{S}$ and b_2 in $\mathbf{S}^\perp$ so that $b_1 + b_2 = b = (1, 1, 1, 1)$.

3.5 The Fast Fourier Transform

The Fourier series is linear algebra in infinite dimensions. The "vectors" are functions $f(x)$; they are projected onto the sines and cosines; that produces the Fourier coefficients a_k and b_k. From this infinite sequence of sines and cosines, multiplied by a_k and b_k, we can reconstruct $f(x)$. That is the classical case, which Fourier dreamt about, but in actual calculations it is the ***discrete Fourier transform*** that we compute. Fourier still lives, but in finite dimensions.

This is pure linear algebra, based on orthogonality. The input is a sequence of numbers $y_0, \ldots, y_{n-1}$, instead of a function $f(x)$. The output $c_0, \ldots, c_{n-1}$ has the same length n. The relation between y and c is linear, so it must be given by a matrix. This is the ***Fourier matrix*** F, and the whole technology of digital signal processing depends on it. The Fourier matrix has remarkable properties.

Signals are digitized, whether they come from speech or images or sonar or TV (or even oil exploration). The signals are transformed by the matrix F, and later they can be transformed back—to reconstruct. What is crucially important is that F and F^{-1} can be quick:

$$\boxed{F^{-1} \textit{ must be simple. The multiplications by } F \textit{ and } F^{-1} \textit{ must be fast.}}$$

Those are both true. F^{-1} has been known for years, and it looks just like F. In fact, F is symmetric and orthogonal (apart from a factor $\sqrt{n}$), and it has only one drawback: Its entries are ***complex numbers***. That is a small price to pay, and we pay it below. The difficulties are minimized by the fact that *all entries of F and F^{-1} tare powers of a single number w*. That number has $w^n = 1$.